Gaviria, Jaime Andrés; López-Ríos, Víctor Ignacio

Locally D-Optimal Designs with Heteroscedasticity: A Comparison between Two Methodologies

Locally D-Optimal Designs with Heteroscedasticity: A Comparison between Two Methodologies

Diseños D-óptimos locales con heterocedasticidad: una comparación entre dos metodologías

Jaime Andrés Gaviria[a], Víctor Ignacio López-Ríos[b]

Escuela de Estadística, Facultad de Ciencias, Universidad Nacional de Colombia, Medellín, Colombia

Abstract

The classic theory of optimal experimental designs assumes that the errors of the model are independent and have a normal distribution with constant variance. However, the assumption of homogeneity of variance is not always satisfied. For example when the variability of the response is a function of the mean, it is probably that a heterogeneity model be more adequate than a homogeneous one. To solve this problem there are two methods: The first one consists of incorporating a function which models the error variance in the model, the second one is to apply some of the Box-Cox transformations to both sides on the nonlinear regression model to achieve a homoscedastic model (Carroll & Ruppert 1988, Chapter 4). In both cases it is possible to find the optimal design but the problem becomes more complex because it is necessary to find an expression for the Fisher information matrix of the model. In this paper we present the two mentioned methodologies for the D-optimality criteria and we show a result which is useful to find D-optimal designs for heteroscedastic models when the variance of the response is a function of the mean. Then we apply both methods with an example, where the model is nonlinear and the variance is not constant. Finally we find the D-optimal designs with each methodology, calculate the efficiencies and evaluate the goodness of fit of the obtained designs via simulations.

Key words: D-efficiency, D-optimal design, Box-Cox transformations, Heteroscedasticity.

---

[a]MSc in Statistics. E-mail: jagaviriab@unal.edu.co

[b]Associate professor. E-mail: vilopez@unal.edu.co

Resumen

La teoría clásica de los diseños experimentales óptimos supone que los errores del modelo son independientes y tienen una distribución normal con varianza constante. Sin embargo, el supuesto de homogeneidad de varianza no siempre se satisface. Por ejemplo, cuando la variabilidad de la respuesta es una función de la media, es probable que un modelo heterocedástico sea más adecuado que uno homogéneo. Para solucionar este problema hay dos métodos: el primero consiste en incorporar una función que modele la varianza del error en el modelo; el segundo consiste en aplicar alguna de las transformaciones de Box-Cox en el modelo de regresión no lineal (Carroll & Ruppert 1988, Capítulo 4). En ambos casos es posible hallar el diseño óptimo, pero el problema se vuelve más complejo porque es necesario encontrar una expresión de la matriz de información de Fisher del modelo. En este artículo se presentan las dos metodologías mencionadas para el criterio D-optimalidad y se muestra un resultado que es útil para encontrar diseños D-óptimos para modelos heterocedásticos cuando la varianza de la respuesta es una función de la media. Luego, se aplican ambos métodos en un ejemplo donde el modelo es no lineal y la varianza no constante. Finalmente se encuentra el diseño D-óptimo con cada metodología, se calculan las eficiencias y se evalúa la bondad del ajuste de los diseños obtenidos a través de simulaciones.

Palabras clave: D-eficiencia, Diseños D-óptimos, heterocedasticidad, transformación de Box-Cox.

1. Introduction

The optimal experimental designs are a tool that allows the researcher to know which factor levels should be experimented in order to obtain a best estimate of the parameters of the model with certain statistical criterion. One of the most popular criteria is the D-optimality which involves finding the design that minimizes the generalized variance of the parameter vector. The design depends on a regression model (1) that relates the response variable Y with the independent variable x

$$Y = \eta(x, \boldsymbol{\beta}) + \epsilon \tag{1}$$

with $\eta(x, \boldsymbol{\beta})$ a linear or nonlinear function of the parameter vector $\boldsymbol{\beta}$ and x.

Besides, if the researcher has the possibility to run N observations of the model (1), then there are the following assumptions:

1. the error components ϵ_i, for $i = 1, 2, \ldots, N$, are independent and

2. have a normal distribution with constant variance σ^2.

For more information about the classic theory of optimal designs see Kiefer (1959), O'Brien & Funk (2003), Atkinson, Donev & Tobias (2007, Chapter 9), López & Ramos (2007). However, in practice there are cases where the homogeneity assumption is not satisfied. For example when the variance of the response

is a function of the mean, it can increase or decrease depending of the structure of the variance. The issue of heteroscedasticity in nonlinear regression models is discussed in detail in Seber & Wild (1989, pp. 68-72). Basically there are two methodologies to handle this problem. The first one is to apply some of the Box-Cox transformations to the model (1) with an appropriate λ_1 that stabilizes the variance of the errors. We identify the transformed model like model A:

$$Y_i^{\lambda_1} = \eta(x_i, \boldsymbol{\beta})^{\lambda_1} + \epsilon_i^* \tag{2}$$

where is assumed that the new errors ϵ_i^* have a normal distribution with constant variance.

The second model, which we identify as model B, consists of incorporating the variance structure of the errors in the model as follows:

$$Y_i = \eta(x_i, \boldsymbol{\beta}) + \epsilon_i \tag{3}$$

where the errors ϵ_i are independent $N(0, \sigma^2(\eta(x_i, \boldsymbol{\beta}))^{\lambda_2})$, with λ_2 an adequate power parameter that models the variance of the errors.

As Seber & Wild (1989) emphasize, the difference between models A and B is "that model A transforms so that $y^{(\lambda_1)}$ has a different distribution from y as well as having a homogeneous variance structure, while model B models the variance heterogeneity but leaves the distribution of y unchanged". Also, the authors affirm that model B has often been preferred to model A when the deterministic function is linear, whereas models like A have been preferred in nonlinear models.

Now, in the context of optimal designs when the model has heteroscedasticity, the problem to find D-optimal designs is more complicated than in the homogeneous case, because the D-optimality criterion maximizes the determinant of the Fisher information matrix of the model and the expression of this matrix changes when the variance is not constant. Because the information matrix depends of the model used, the two methodologies mentioned before for handling of heteroscedasticity are traditionally applied in separate ways. For example, Atkinson & Cook (1997) apply some of the Box-Cox transformations that makes the transformed model be linear with a constant variance and then they find local and Bayesian D-optimal designs to several models. On the other hand, in the case of linear models, Atkinson & Cook (1995) find local D-optimal designs for heteroscedastic linear models for various structures of variance, one of them is when the logarithm of variance is a linear function of the independent variable. Other authors have worked with nonlinear models, see for example Dette & Wong (1999).

In this paper we compare the methodologies mentioned above, analyze the structure of the information matrix and we find the D-optimal design for a specific model. Finally we compare the designs obtained through the D- efficiency. This paper is divided in four sections. In section 2 we present a brief summary of both methodologies for the D-optimality criterion and show a result which is useful to find D-optimal designs for heteroscedastic models when the variance of the response is a function of the mean (we omit the proof due to length constraints). In section 3 we illustrate both methods with an example and we compare results using the D-efficiency of each design. Then, we simulate observations of the model

for each design and we calculate the relative error and mean square error. Finally, in section 4 we present some conclusions, discussions and suggestions.

2. Methodologies

Starting with a regression model of the form (1) with the usual assumptions, the problem of optimal designs consists to find the levels of the x factor where the researcher should experiment to obtain a best estimate of the parameters of the model under certain statistical criterion. In this paper we focus on the D-optimality criterion, which finds the design that minimizes the generalized variance of the parameter vector (Atkinson et al. 2007, pp. 135). More precisely, a design ξ is defined as a measure of probability with finite support denoted by:

$$\xi = \left[\begin{array}{cccc} x_1 & x_2 & \cdots & x_n \\ w_1 & w_2 & \cdots & w_n \end{array} \right] \tag{4}$$

where n is the number of support points, $x_1, x_2, \ldots, x_n$ are the support points of the design with associated weights $w_i \geq 0$ and such that $\sum_{i=1}^{n} w_i = 1$ (O'Brien & Funk 2003). If the weight w_i is any number between 0 and 1, the design ξ is known as a continuous design. However in practice all designs are exact. This means that the weights w_i are associated with the frequency of the support points (Atkinson et al. 2007, pp. 120).

Now, the main problem of optimal designs is to find a design ξ over a compact region χ, that maximizes a functional of the information matrix $M(\xi)$. This matrix plays an important role in the theory of optimal experimental designs. The structure of this matrix depends on the linear nature of the model and on the assumptions about the errors. When the variance of the errors is constant, this matrix has a known expression, see for example López & Ramos (2007). However, in the case of heteroscedastic models this expression is more complex and depends of the methodology applied. So, in the next two sections we analyze the structure of the Fisher information matrix with each of the two methodologies mentioned before.

2.1. Variance Modelling

When the variance of the error is not constant, one way to solve this problem is to find an adequate function which models the error variance and incorporate it in the regression model. There are many ways to do this, see for example Huet, Bouvier, Poursat & Jolivet (2004, pp. 65) and Seber & Wild (1989, pp. 68-72). One form is when the variance of the response is a power function of the mean:

$$Y_i = \eta(x_i, \boldsymbol{\beta}) + \epsilon_i, \text{ with } var(\epsilon_i) = \sigma^2(\eta(x_i, \boldsymbol{\beta}))^{2\tau} \tag{5}$$

where σ^2 is the constant variance, τ is an unknown parameter and it should be estimated. The model (5) with variance structure is known as the power of the mean variance model (Ritz & Streibig 2008, pp. 74).

Now, some authors have worked the problem to find *D*-optimal designs modeling the variance. For example Dette & Wong (1999) find *D*-optimal designs for the Michaelis-Menten model when the variance is a function of the mean and Atkinson & Cook (1995) find *D*-optimal designs for heteroscedastic linear models. The following result is taken from Downing, Fedorov & Leonov (2001); they show the expression of the information matrix for a more general model than the power of the mean variance model (5):

$$Y = \eta(x, \boldsymbol{\theta}) + \epsilon, \text{ with } var(\epsilon) = S(x, \boldsymbol{\theta}) \tag{6}$$

where $\boldsymbol{\theta}$ is the parameter vector and it can include the parameters of the deterministic function η and those of the function $S(x, \boldsymbol{\theta})$ as a positive function used to model the variance of the error. Observe that the power of the mean variance model (5) is a nested model of the more general model (6). In this case the parameter vector $\boldsymbol{\theta}$ includes all the possible parameters of the model: $\boldsymbol{\beta}, \tau$ and σ^2. So, results about the general model (for instance the next theorem)can be applied in particular for the power of the mean variance model.

Theorem 1. *Information Matrix.*

Let Y with normal distribution, with expected mean $E[Y|x] = \eta(x, \boldsymbol{\theta})$ and variance $Var[Y|x] = S(x, \boldsymbol{\theta})$, where $S(x, \boldsymbol{\theta}) > 0$ is a positive function, $\boldsymbol{\theta}_{q \times q}$ is the parameter vector and χ a compact set. If the N observations $\{y_i, x_i\}_{i=1}^{N}$ are independent, then the Fisher information matrix for the approximate design ξ over the regression design χ is

$$\boldsymbol{M}(\xi, \boldsymbol{\theta})_{q \times q} = \int_{\chi} \boldsymbol{I}(x, \boldsymbol{\theta}) d\xi(x) \tag{7}$$

where

$$I(x, \boldsymbol{\theta})_{q \times q} = \frac{1}{S(x, \boldsymbol{\theta})} \frac{\partial \eta(x, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \eta(x, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} + \frac{1}{2} \frac{1}{S(x, \boldsymbol{\theta})^2} \frac{\partial S(x, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial S(x, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} \tag{8}$$

This theorem is the main tool of this methodology, because it allows the researcher many ways of modelling the variance and incorporate it in the model.

Corollary 1. *For the power of the mean variance model given in (1), where the errors are independent and have normal distribution with mean zero and variance $var(\epsilon_i) = \sigma^2(\eta(x_i, \boldsymbol{\beta}))^{2\tau}$ with $\boldsymbol{\beta}, \tau$ and σ^2 parameters, the information matrix is given by*

$$\boldsymbol{M}(\xi, \boldsymbol{\theta}) = \boldsymbol{U}\boldsymbol{W}\boldsymbol{U}^T + \boldsymbol{V}\boldsymbol{W}\boldsymbol{V}^T \tag{9}$$

where

$$\boldsymbol{U}_{(p+2) \times n} = (\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n) \quad \boldsymbol{V}_{(p+2) \times n} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n) \tag{10}$$

$$\boldsymbol{W} = \begin{pmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & & \\ \vdots & & \ddots & \\ 0 & & & w_r \end{pmatrix} \tag{11}$$

and for $i = 1, 2, \ldots, n$:

$$\boldsymbol{u}_i = \left(\frac{1}{\sigma \eta(x_i, \boldsymbol{\beta})^\tau} \frac{\partial \eta(x_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T}, 0.0 \right)^T_{(p+2) \times 1} \tag{12}$$

$$\boldsymbol{v}_i = \left(\frac{\sqrt{2}\tau}{\eta(x_i, \boldsymbol{\beta})} \frac{\partial \eta(x_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T}, \sqrt{2} \log \eta(x_i, \beta), \frac{1}{\sqrt{2}\sigma^2} \right)^T_{(p+2) \times 1} \tag{13}$$

This result is the key at the construction of D-optimal designs and can be implemented computationally to obtain the designs. We will illustrate the use of this corollary with an application in the next section. But before we need the following important result which is one of the equivalence theorems. This theorem allows to verify if the obtained design is in fact the optimal design (Kiefer & Wolfowitz 1960)

Theorem 2. *D-optimality equivalence theorem.*

Let $\boldsymbol{M}(\xi, \boldsymbol{\theta})_{q \times q}$ the information matrix of the design ξ positive, $\Psi(\xi, \boldsymbol{\theta}) = \log |\boldsymbol{M}(\xi, \boldsymbol{\theta})|$ the D-optimality criterion and χ a compact set. Then the design ξ^* is D-optimal if the directional derivative of ϕ in ξ^* on the direction of ξ_x holds

$$\phi(M(\xi^*, \boldsymbol{\theta}), M(\xi_x, \boldsymbol{\theta})) \leq 0 \quad \forall x \in \chi \tag{14}$$

where $\phi(\boldsymbol{M}(\xi^*, \boldsymbol{\theta}), M(\xi_x, \boldsymbol{\theta})) = Tr(\boldsymbol{M}(\xi_x)\boldsymbol{M}^{-1}(\xi^*)) - q$ and ξ_x is the design that puts all probability in x. Also, $\phi(\boldsymbol{M}(\xi^*), \boldsymbol{M}(\xi_x)) = 0$ at the support points of design ξ^*.

This result is useful to verify the D-optimality of a design ξ^*, because one can plot the directional derivative $\phi(M(\xi^*, \theta), M(\xi_x, \theta))$ over $x \in \chi$ and to check that this function at most zero over all experimental region (χ) and also that in the support points of the design, the equality holds.

2.2. Transformation of the Model

The second methodology consists of applying an adequate transformation on the model to obtain constant variance. We focus on the Box-Cox transformations, which are given by Box & Cox (1964).

$$y^{(\lambda)} = \begin{cases} \frac{y^\lambda - 1}{\lambda} & \text{for } \lambda \neq 0 \\ \log y & \text{for } \lambda = 0 \end{cases} \tag{15}$$

The value of the parameter λ usually is unknown, but in some cases it can be assessed depending on the response. For instance, if the response is a volume, the appropriate transformation can be the cube root ($\lambda = 1/3$) and the square root if the response corresponds to count data (Atkinson & Cook 1997).

Now, Atkinson & Cook (1997) find D-optimal designs when a Box-Cox transformation is applied, the resulting model is linear

$$\boldsymbol{Y}^{(\lambda)} = f^T(x)\boldsymbol{\beta} + \epsilon^* \tag{16}$$

and the errors have normal distribution with constant variance $\epsilon^* \sim N(0, \sigma^2)$.

However, as illustrated in the example and since the original model is nonlinear, we must find some appropriate λ such that when we apply the transformation to both sides of the model, the transformed model is linear in the parameters. It is important to observe that in our case, the parameter λ will be known, which is an advantage, because we do not need to estimate this parameter. However, when the parameter λ is unknown, it is possible to find the design, see for example Atkinson (2003) for more details.

Then, the authors show that the information matrix over the design region for the transformed model is (see the details in Atkinson & Cook 1997)

$$\boldsymbol{M}(\xi, \theta) = \int_\chi \boldsymbol{I}(\theta)\xi(dx) \tag{17}$$

where the symmetric matrix $I(\theta)$ is given by

$$
\begin{aligned}
I(\theta) &= -E\left[\frac{\partial^2 \log f(\boldsymbol{Y}_i | x_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2}\right] \\
&= \begin{pmatrix} ff^T & 0 & -\frac{fE(\dot{Y}^{(\lambda)})}{\sigma^2} \\ 0 & \frac{1}{2\sigma^4} & -\frac{E(\epsilon^* \dot{Y}^{(\lambda)})}{\sigma^4} \\ -\frac{fE(\dot{Y}^{(\lambda)})}{\sigma^2} & -\frac{E(\epsilon^* \dot{Y}^{(\lambda)})}{\sigma^4} & \frac{E(\dot{Y}^{(\lambda)})^2 + E(\epsilon^* \ddot{Y}^{(\lambda)})}{\sigma^2} \end{pmatrix}
\end{aligned} \tag{18}
$$

with $\epsilon^* = Y^{(\lambda)} - f^T(x)\beta$, $f = f(x)$ and $\dot{Y}^{(\lambda)}, \ddot{Y}^{(\lambda)}$ denote the first and second derivative respect to λ and are given by:

$$\dot{Y}^{(\lambda)} = \frac{Y^\lambda \log Y^\lambda - Y^\lambda + 1}{\lambda^2} \quad \text{and} \tag{19}$$

$$\ddot{Y}^{(\lambda)} = \frac{Y^\lambda (\log Y^\lambda - 1)^2 + Y^\lambda - 2}{\lambda^3} \tag{20}$$

However, these expressions have to be approximated using first-order Taylor approximations, since the expected values can not be calculated exactly. Finally, once the design is found using the above expressions, is necessary verify the *D*-optimality of the design using a similar result of the equivalence theorem 2.

3. Example

In Section 2 we described the two methodologies commonly used to handle the heteroscedasticity of a model. Now we illustrate these methods with one example.

3.1. PCB Model

The example consists of a study realized in 1972 in Lake Cayuga, New York, where the concentrations of *Polychlorinated biphenyls* (PCB) were made in a group

of 28 trout at several ages in years. "The ages of the fish were accurately known because the fish are annually stocked as yearlings and distinctly marked as to year class" (Bates & Watts 1988, pp. 267–268). The data taken from Bates & Watts (1988), are shown in the table 1 and the scatter plot is shown in figure 1.

TABLE 1: Lake Cayuga data.

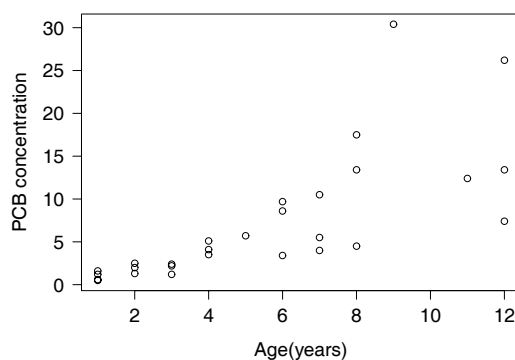| Age	1.00	1.00	1.00	1.00	2.00	2.00	2.00
Concentration	0.60	1.60	0.50	1.20	2.00	1.30	2.50
Age	3.00	3.00	3.00	4.00	4.00	4.00	5.00
Concentration	2.20	2.40	1.20	3.50	4.10	5.10	5.70
Age	6.00	6.00	6.00	7.00	7.00	7.00	8.00
Concentration	3.40	9.70	8.60	4.00	5.50	10.50	17.50
Age	8.00	8.00	9.00	11.00	12.00	12.00	12.00
Concentration	13.40	4.50	30.40	12.40	13.40	26.20	7.40

FIGURE 1: Scatter plot of Lake Cayuga data.

The plot of the data shows that the concentration of *Polychlorinated biphenyls* (PCB) increases when the age of the trout does. Also, the relationship between the variables clearly is not linear, so we propose to fit the nonlinear model:

$$Y = \beta_1 e^{\beta_2 x} + \epsilon \tag{21}$$

with β_1, β_2 are unknown parameters to be estimated.

Now, we are going to find the D optimal design for this model with the two methodologies described above. Because our designs are local, we use the data only with the purpose to have a good local value of the parameter vector.

3.1.1. Variance Modelling

First, we apply the methodology consisting on modelling the variance of the errors with an appropriate function. In figure 1, we see that the variability of the concentration increases as a power function of the mean, so we propose to fit the model (21) with variance structure (5)

$$Y = \beta_1 e^{\beta_2 x} + \epsilon, \text{ where } \epsilon \sim N(0, \sigma^2 (\beta_1 e^{\beta_2 x})^{2\tau}) \tag{22}$$

with τ an unknown parameter to be estimated.

Now, we fit the model (22) in R Development Core Team (2013) and we used the *gnls* function for the generalized nonlinear least squares method. The results of the estimation are showed in table 2.

TABLE 2: Generalized nonlinear least squares estimation.

Parameter	Estimation
β_1	0.91
β_2	0.31
τ	1.19
σ	0.34

Next, we perform the likelihood ratio test to determine if the model with variance structure (22) is better than the model with constant variance (21). The results of the test are showed in table 3 (the model 1 corresponds to the model with variance structure (22) and the model 2 to the model with constant variance (21)).

TABLE 3: ANOVA for the likelihood ratio test.

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
(22)	1	4	134.5534	139.8822	-63.27671			
(21)	2	3	178.8002	182.7968	-86.40008	1 vs 2	46.24674	<.0001

The conclusion from this test that is the parameter $\tau \neq 0$, e.g. the model with variance structure (22) is better than the model with constant variance (21) with a signification level of 1%.

3.1.2. *D*-Optimal Design

Now, we find the *D*-optimal design for the model with variance structure (22). Because we work with local designs, we use the estimation of the parameters obtained previously like the local value for θ; that is, we use the local value $\boldsymbol{\theta}_0 = (\beta_1, \beta_2, \tau, \sigma) = (0.91, 0.31, 1.19, 0.34)$. Then we implement the corollary 1 through an algorithm in R Development Core Team (2013) and minimize $-\log(|\boldsymbol{M}(\xi, \boldsymbol{\theta})|)$. In this optimization problem we use the function *nlminb* over the experimental region (χ). The local D-optimal design obtained is shown in table 4 and is denoted by ξ_D. The x_i are the support points of the design and the w_i the weights. As we can see, even though the model with variance structure (22) has four parameters to be estimated, the design consists only of two points, which are the extreme points of the regression range $\chi = [1, 12]$. In this sense, if we could repeat the experiment and our objective are to estimate the parameters with minimum variance, then we measure the *Polychlorinated biphenyls* concentration in trout with ages of one and twelve and with equal number of replicates.

Then we check that the obtained design ξ_D is *D*-optimal. With this in mind, by the *D*-optimality equivalence in theorem 2, we must verify that the directional derivative of Ψ at ξ_D in the direction of the design that puts all mass at x, ξ_x,

TABLE 4: Local D-optimal design ξ_D to the model (22).

x_i	1.00	12.00
w_i	0.50	0.50

satisfies

$$\phi(\boldsymbol{M}(\xi_D, \boldsymbol{\theta}), \boldsymbol{M}(\xi_x, \boldsymbol{\theta})) = Tr(\boldsymbol{M}(\xi_x)\boldsymbol{M}^{-1}(\xi_D)) - 4 \leq 0 \qquad (23)$$

$\forall x \in \chi = [1, 12]$. As we can see in figure 2, this condition holds and the derivative equals zero at the support points, so the design ξ_D is indeed D-optimal.
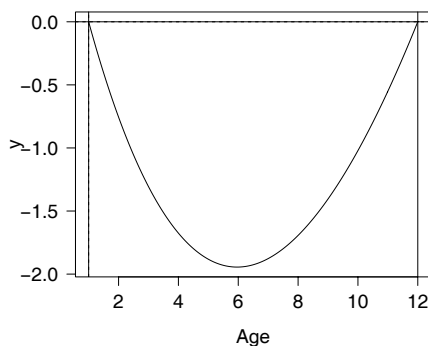

FIGURE 2: Plot of the directional derivative.

3.1.3. Simulations

We simulate 1,000 times 28 observations of the model with variance structure

$$Y_i = \beta_1 e^{\beta_2 x_i} + \epsilon_i, \ \epsilon_i \sim N(0, \sigma^2(\beta_1 e^{\beta_2 x_i})^{2\tau}), \ \text{for } i = 1, 2, \ldots, 28 \qquad (24)$$

taking the values of x_i like the support points of the design ξ_D. Then we simulate the errors $\epsilon_i \sim N(0, \sigma^2(\beta_1 e^{\beta_2})^{2\tau})$ for $i = 1, 2, \ldots, 28$; use the estimations obtained in table 2 like the values of the parameters and with the model (24), we calculate the response $y_i's$. Then, with these simulated data, we obtain the estimated parameter $\hat{\theta}$ and calculate the relative and mean square error (RE and MSE respectively). We repeat this process 1,000 times and summarize it in table 5, showing the descriptive measures for both errors. This table shows the mean, median, range and standard deviation for the MSE of each parameter of the model (24) and the relative error in percentage $RE(\theta) \times 100\%$. For the parameter vector θ we propose an overall discrepancy measure, ODM, defined as $ODM(\hat{\theta}) = ||\theta - \hat{\theta}||^2$. From this table, we see that the central tendency measures for the MSE are small as the variability between the simulations. Also, the mean and median for the RE are very close to 10%. In general, all these measures indicate that the local design ξ_D provides good parameter estimates, even though the design only has two experimental points and the model four parameters.

TABLE 5: Simulations with variance modelling (Std denotes the Standard deviation).

	$MSE(\beta_1)$	$MSE(\beta_2)$	$MSE(\tau)$	$MSE(\sigma)$	$ODM(\theta)$	$RE(\theta)\%$
Mean	9.08e-03	3.13e-04	1.08e-02	5.97e-03	2.61e-02	9.31
Median	4.28e-03	1.50e-04	5.40e-03	2.68e-03	1.86e-02	8.71
Range	9.60e-02	5.78e-03	1.15e-01	1.09e-01	2.01e-01	27.60
Std	1.28e-02	4.61e-04	1.45e-02	8.84e-03	2.56e-02	4.45

3.1.4. Efficiencies

Finally, we show the robustness of the design ξ_D with respect to the choice of the local value θ_0, through the D-efficiency of any design ξ:

$$D_{\text{eff}} = \left(\frac{|\boldsymbol{M}(\xi)|}{|\boldsymbol{M}(\xi_D)|} \right)^{1/p} \tag{25}$$

where p is the number of parameters of the model and $\boldsymbol{M}(\xi)$ denotes the information matrix of the design, where ξ is another design obtained with another local values of parameter vector. With this in mind, we perturb each one of the four parameters of the model (22) in a percentage Δ:

$$\theta_i \pm \Delta \times \theta_i \tag{26}$$

Since the model has four parameters and each one can be perturbed at left, at right or not be perturbed; it is clear that the total number of perturbations is $3^4 = 81$. Then each one of these perturbations will give us a design ξ and with (25) we calculate how far we are of the local D-optimal design. Then for a fixed $\Delta = 0.6$ (we could used another), we obtain 81 designs and for each one we calculate the respective D-efficiency. However, because most of these designs were equal to the two point design ξ_D, we only show in table 9 (see the appendix) the support points, the weights and the D-efficiency of the 36 designs that were different to the optimal. Figure 3 summarizes the results of the efficiencies and shows that the design ξ_D is robust respect the choice of the local value θ_0, because the D-efficiencies are high (at least 0.80).

3.2. Transformation of the Model

Previously we apply the first methodology of variance modelling and find the local D-optimal design. Now we use the second methodology, that consists on applying an adequate transformation on the model. As we described in section 2.2, this transformation should be such that the transformed model is linear and homoscedastic. In this case as the model (1) is exponential, the appropriate Box-Cox transformation consists on applying logarithm to both sides:

$$\log Y = \log \beta_1 + \beta_2 x + \epsilon^* \tag{27}$$

or equivalently in the form:

$$Y^* = \beta_1^* + \beta_2 x + \epsilon^* \tag{28}$$
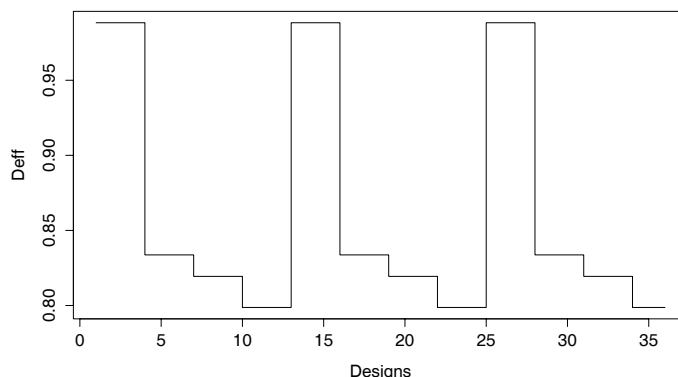
FIGURE 3: D-efficiencies perturbing θ in 60%.

where $Y^* = \log Y$, $\beta_1^* = \log \beta_1$ and the new errors ϵ^* are normal with constant variance. Then we fitted the linear model (28) and we obtained the estimations $\hat{\boldsymbol{\beta}}^* = (0.03, 0.26)^T$, $\hat{\sigma} = 0.57$, and then $\hat{\boldsymbol{\beta}} = (e^{0.03}, 0.26)^T = (1.03, 0.26)^T$. Finally, we implemented an algorithm in R Development Core Team (2013) to find the information matrix with $\lambda = 0$ and to obtain the design that minimizes $-\log \boldsymbol{M}(\xi, \boldsymbol{\theta})$ over $\chi = [1, 12]$. The resulting design in table 6, shows that in this case the design is the same obtained with first methodology. However, we have to point out that despite that the resulting design is the same with both methodologies, it is attributed to the fact that with each method we used the best local value for the parameter $\boldsymbol{\theta}$ and as we saw when we calculate the D-efficiencies, the design can have three support points depending on the local value used.

TABLE 6: D-optimal design to the model (27).

i	1	2
x_i	1.00	12.00
w_i	0.50	0.50

3.2.1. Simulations

Analogously to the first methodology, we simulated 28 observations of the model (28). The results of the $1,000$ simulations are summarized in table 7. This is similar to the table 5 and shows the mean, median, range and standard deviation for the MSE of each parameter of the model (27) and relative error in percentage $RE(\theta) \times 100\%$. For the parameter vector $\boldsymbol{\theta}$ we propose a measure defined as $ODM(\hat{\boldsymbol{\theta}}) = ||\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}||^2$, which is a kind of square distance between the estimated parameter and the original. The conclusions from these results are similar as the obtained with the first methodology, although when we compare the measures for the relative error (RE), is noteworthy that all the descriptive measures are almost three times the correspondent to the first methodology. But in general, all these measures indicate that the local design ξ_D fits well the model.

Table 7: Simulations for the logarithmic transformation model (Std denotes the Standard deviation).

	$MSE(\beta_1)$	$MSE(\beta_2)$	$MSE(\sigma)$	$ODM(\theta)$	$RE(\theta)\%$
Mean	1.80e-01	9.57e-04	1.27e-02	1.93e-01	34.0
Median	1.39e-01	6.48e-04	8.43e-03	1.53e-01	32.4
Range	1.18e+00	8.65e-03	9.34e-02	1.18e+00	84.0
Std	1.55e-01	1.00e-03	1.36e-02	1.53e-01	13.1

3.2.2. Efficiencies

Finally, we obtain the D-efficiencies following the same procedure described in section 3.1.4. In this case because we perturb three parameters: β_1, β_2 and σ, we only have $3^3 = 27$ combinations (the parameter $\lambda = 0$). But again most of all these designs were equal to the D-optimal design, so we only show in table 8 the six designs that correspond to a perturbation $\Delta = 60\%$ and were different to the optimal. In this table we use the symbols $-$, $+$ or 0 to indicate the specific combinations of the parameters.

For instance, the first design is obtained when we disturb 60% to the left ($-$) the parameters β_1 and σ and we do not perturb (0) the parameter β_2. Then the support points for this design are 1, 6.5 and 12 and the D-efficiency of the design is 0.93. It indicates that if we use this design instead of the unperturbed D-optimal design, we would need around 7% more observations to obtain the same efficiency that the D-optimal. Even more, it is remarkable that all six designs have exactly 3 support points: The extremes of the interval $[1, 12]$ and the middle point 6.5. The only difference between these designs is the weight (in parentheses with two decimal places) and the D-efficiency, that can be 0.89 or 0.93, but in both cases it is high, so we can conclude that the D-optimal design is robust respect to the choice of the local value θ_0.

Table 8: Support points, weights and D-efficiencies perturbing 60% to left ($-$), right ($+$) or not (0).

Design	β_1	β_2	σ	x_1	x_2	x_3	D_{eff}
1	$-$	0	$-$	1(0.40)	6.5(0.20)	12(0.40)	0.93
2	0	0	$-$	1(0.40)	6.5(0.20)	12(0.40)	0.93
3	$+$	0	$-$	1(0.40)	6.5(0.20)	12(0.40)	0.93
4	$-$	$+$	$-$	1(0.36)	6.5(0.28)	12(0.36)	0.89
5	0	$+$	$-$	1(0.36)	6.5(0.28)	12(0.36)	0.89
6	$+$	$+$	$-$	1(0.36)	6.5(0.28)	12(0.36)	0.89

4. Conclusions

We have presented a brief summary of two methodologies that can be implemented to find D-optimal designs when the model under study presents heteroscedasticity. In both cases the main problem is to find an expression for the Fisher information matrix of the model. We have illustrated both methods with

Lake Cayuga data from which clearly do not have constant variance. However, there is an important difference between the methods that applied: The variance modelling methodology has the assumption that the errors of the original model has a normal distribution. However, the second methodology only requires a normal distribution for the transformed model, not the original. This can be an advantage of this methodology compared with variance modelling.

Under both methods, we find the same D-optimal design with two support points and with equal weights. But this fact is attributed only to the local values used in an independent way, that in this case were the estimations of the parameter vector using the data. Because the optimal design is local, we determine the robustness of this design with each methodology by disturbing the parameters of the corresponding model and calculating the D-efficiency of the obtained designs. In both cases, the efficiencies were high indicating that the D-optimal design is a robust design respect the choice of the local value θ_0. Also, with each methodology we simulate $1,000$ observations of the model and calculate some descriptive measures for the relative and mean square errors. The results were similar. The only important difference is that measures for the relative errors of the second methodology were almost three times the correspondent to the first methodology. We cannot conclude which methodology is better because each one has its pros and shortcoming, with the example we obtained similar results.

Finally, we want to point out that we have not study two potential problems: First, the problem of heteroscedasticity for G optimality criterion and second, the problem of nonnormality (for D-optimality or not). Respect to the former, further work includes finding optimal designs for heteroscedastic models with another optimality criteria different to D-optimality. For instance, Wong & Cook (1993) have worked with G-optimal designs with linear models when the variance of the errors is incorporated in the model. With non normality, we did not find too many published papers, so this can be an interesting problem to work. Finally we have found local designs, but other option is to use the Bayesian approach.

TABLE 9: *D*-efficiencies, support points and weights with a 60% of perturbation of the parameter vector: disturb to left (−), to right (+) or do not (0).

ξ_i	β_1	β_2	τ	σ	x_1	x_2	x_3	w_1	w_2	w_3	D_{eff}
1	−	0	−	−	1	5.6	12	0.48	0.02	0.5	0.9883
2	0	0	−	−	1	5.6	12	0.48	0.02	0.5	0.9883
3	+	0	−	−	1	5.6	12	0.48	0.02	0.5	0.9883
4	−	+	−	−	1	8.26	12	0.27	0.28	0.45	0.8337
5	0	+	−	−	1	8.26	12	0.27	0.28	0.45	0.8337
6	+	+	−	−	1	8.26	12	0.27	0.28	0.45	0.8337
7	−	0	+	−	1	4.45	12	0.44	0.3	0.26	0.8194
8	0	0	+	−	1	4.45	12	0.44	0.3	0.26	0.8194
9	+	0	+	−	1	4.45	12	0.44	0.3	0.26	0.8194
10	−	+	+	−	1	3.12	12	0.42	0.34	0.24	0.7987
11	0	+	+	−	1	3.12	12	0.42	0.34	0.24	0.7987
12	+	+	+	−	1	3.12	12	0.42	0.34	0.24	0.7987
13	−	0	−	0	1	5.6	12	0.48	0.02	0.5	0.9883
14	0	0	−	0	1	5.6	12	0.48	0.02	0.5	0.9883
15	+	0	−	0	1	5.6	12	0.48	0.02	0.5	0.9883
16	−	+	−	0	1	8.26	12	0.27	0.28	0.45	0.8337
17	0	+	−	0	1	8.26	12	0.27	0.28	0.45	0.8337
18	+	+	−	0	1	8.26	12	0.27	0.28	0.45	0.8337
19	−	0	+	0	1	4.45	12	0.44	0.3	0.26	0.8194
20	0	0	+	0	1	4.45	12	0.44	0.3	0.26	0.8194
21	+	0	+	0	1	4.45	12	0.44	0.3	0.26	0.8194
22	−	+	+	0	1	3.12	12	0.42	0.34	0.24	0.7987
23	0	+	+	0	1	3.12	12	0.42	0.34	0.24	0.7987
24	+	+	+	0	1	3.12	12	0.42	0.34	0.24	0.7987
25	−	0	−	+	1	5.6	12	0.48	0.02	0.5	0.9883
26	0	0	−	+	1	5.6	12	0.48	0.02	0.5	0.9883
27	+	0	−	+	1	5.6	12	0.48	0.02	0.5	0.9883
28	−	+	−	+	1	8.26	12	0.27	0.28	0.45	0.8337
29	0	+	−	+	1	8.26	12	0.27	0.28	0.45	0.8337
30	+	+	−	+	1	8.26	12	0.27	0.28	0.45	0.8337
31	−	0	+	+	1	4.45	12	0.44	0.3	0.26	0.8194
32	0	0	+	+	1	4.45	12	0.44	0.3	0.26	0.8194
33	+	0	+	+	1	4.45	12	0.44	0.3	0.26	0.8194
34	−	+	+	+	1	3.12	12	0.42	0.34	0.24	0.7987
35	0	+	+	+	1	3.12	12	0.42	0.34	0.24	0.7987
36	+	+	+	+	1	3.12	12	0.42	0.34	0.24	0.7987

References

Atkinson, A. C. (2003), 'Horwitz's rule, transforming both sides and the design of experiments for mechanistic models', *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **52**(3), pp. 261–278.

Atkinson, A. C. & Cook, R. D. (1995), 'D-optimum designs for heteroscedastic linear models', *Journal of the American Statistical Association* **90**(429), 204–212.

Atkinson, A. C. & Cook, R. D. (1997), 'Designing for a response transformation parameter', *Journal of the Royal Statistical Society. Series B (Methodological)* **59**(1), 111–124.

Atkinson, A. C., Donev, A. N. & Tobias, R. D. (2007), *Optimum Experimental Designs with SAS*, Oxford Science Publications, New York.

Bates, D. M. & Watts, D. G. (1988), *Nonlinear Regression Analysis and its Applications*, John Wiley and Sons, New York.

Box, G. E. P. & Cox, D. R. (1964), 'An analysis of transformations', *Journal of the Royal Statistical Society. Series B (Methodological)* **26**(2), pp. 211–252.

Carroll, R. & Ruppert, D. (1988), *Transformation and Weighting in Regression, Chapter 4*, Taylor & Francis.

Dette, H. & Wong, K. (1999), 'Optimal Designs When the Variance Is a Function of the Mean', *Biometrics* **55**(3), 925–929.

Downing, D., Fedorov, V. & Leonov, S. (2001), *Extracting Information from the Variance Function: Optimal Design*, Springer, Austria.

Huet, S., Bouvier, A., Poursat, M. & Jolivet, E. (2004), *Statistical Tools for Nonlinear Regression: A Practical Guide With S-PLUS and R Examples* , Springer-Verlag, New York.

Kiefer, J. (1959), 'Optimum experimental designs', *Journal of the Royal Statistical Society. Series B (Methodological)* **21**(2), pp. 272–319.

Kiefer, J. & Wolfowitz, J. (1960), 'The equivalence of two extremum problems', *Canadian Journal of Mathematics* **12**(5), pp. 363–365.

López, V. & Ramos, R. (2007), 'Introducción a los diseños óptimos', *Revista Colombiana de Estadística* **30**(1), 37–51.

O'Brien, T. & Funk, G. M. (2003), 'A gentle introduction to optimal design for regression models', *Journal of the American Statistical Association* **57**(4), 265–267.

R Development Core Team (2013), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
*http://www.R-project.org

Ritz, C. & Streibig, J. (2008), *Nonlinear Regression with R*, Springer, New York.

Seber, G. & Wild, C. (1989), *Nonlinear Regression*, John Wiley, New York.

Wong, W. K. & Cook, R. D. (1993), 'Heteroscedastic G-optimal designs', *Journal of the Royal Statistical Society. Series B (Methodological)* **55**(4), pp. 871–880.